Collaborative Tracking for MRI-Guided Robotic Intervention on the Beating Heart

Y. Zhou[1], E. Yeniaras[1], P. Tsiamyrtzis[2], N. Tsekos[1], and I. Pavlidis[1]

[1] Department of Computer Science, University of Houston, Houston, TX 77024, USA
[2] Department of Statistics, Athens University of Economics, Athens 10434, Greece
{yzhou9,eyeniaras2,nvtsekos,ipavlidis}@uh.edu, pt@aueb.gr

Abstract. Magnetic Resonance Imaging (MRI)-guided robotic interventions for aortic valve repair promise to dramatically reduce time and cost of operations when compared to endoscopically guided (EG) procedures. A challenging issue is real-time and robust tracking of anatomical landmark points. The interventional tool should be constantly adjusted via a closed feedback control loop to avoid harming these points while valve repair is taking place in the beating heart. A Bayesian network of particle filter trackers proves capable to produce real-time, yet robust behavior. The algorithm is extremely flexible and general - more sophisticated behaviors can be produced by simply increasing the cardinality of the tracking network. Experimental results on 16 MRI cine sequences highlight the promise of the method.

1 Introduction

Image guided and robot-assisted (IGRA) surgeries are evolving and may selectively replace endoscopically guided (EG) surgeries in the future. Research is motivated by several IGRA advantages, such as wide field-of-view with planar or volumetric appreciation of the area of operation, minimally invasive processes, and reduction of operating time [1]. A grand challenge in IGRA surgeries is the compensation of tissue motion. This is particularly true in heart operations.

Among the most promising types of IGRA surgeries are the interventional magnetic resonance imaging (MRI) surgeries; distinct advantages include lack of ionizing radiation, a wide range of soft tissue contrast mechanisms, 3D data acquisition, and operator-independent image quality [2]. MRI-guided robotic interventions have a wide range of potential applications [3], including cardiac procedures, such as aortic valve repair [4]. Tracking cardiac motion is a very active field within the cardiovascular MRI community [5]. Nevertheless, a method to estimate the motion of specific anatomical landmarks needed in surgical procedures has not yet been proposed.

The research described in this paper is motivated by the need to develop alternatives to the highly invasive and long (several hours) surgical procedures related to heart valve repairs. An MRI-guided robotic intervention will obviate the need to open the thorax and stop the beating heart, thus potentially completing the surgery within minutes. For this, the motion of specific anatomical

T. Jiang et al. (Eds.): MICCAI 2010, Part III, LNCS 6363, pp. 351–358, 2010.

landmarks should be tracked in real-time to close the feedback control loop. As a result, the robotic interventional tool will be held in place during the procedure without harming healthy tissue and vital structures.

To locate a specific anatomical landmark, one approach is to segment the underlying structures, i.e., accurately determine the boundary of the endocardium and the left ventricle on short and long axis views of the heart. Such approach may entail defining an appropriate optimization scheme to iteratively minimize a cost function, and the computational cost can be significant. For example, in [6], while the proposed boundary segmentation method achieved very high accuracy, the computational time was 10 seconds per slice. In contrast, Yuen et al. [7] achieved real-time performance in Ultrasound images with an extended Kalman filter tracker.

We adopted a particle filtering approach to estimate landmark motion. Particle filtering is a general tracking mechanism [8], free of strong modeling, which can accommodate very efficiently the predict-update loop. A loner particle filter tracker, however, may become unstable in sudden motion or large appearance changes. This is risky in cardiac surgeries. To maintain robustness we propose a collaborative tracking framework, which coordinates multiple particle filter trackers. Some of these trackers may fail when confronting challenging conditions, such as sudden motion or significant appearance changes, but others that are less affected will survive and yield good state estimates. The latter can be used to produce a reliable overall estimate and eventually recover the failed trackers. For the first time, we use a Bayesian network method to decide which trackers fail and which ones survive at each time step.

The methodology is quite elegant, as it provides a unified framework to tackle a wide variety of tracking problems, from the most mundane to the most difficult, by simply adjusting the number of trackers n in the probabilistic network. For easy problems, we typically choose $n = 1$, while for more challenging problems $n > 1$. The algorithm achieves robust and real-time performance, as the experimental results indicate. Specifically, the algorithm's ability to monitor the motion of the apex, the center of the left ventricle, and the aortic annulus has been tested with success on cine long and short axis MRI imagery.

2 Methodology

2.1 Single Particle Filter Tracker ($n = 1$)

The particle filter tracker that we use features 100 particles and performs a single iteration per frame. We denote the motion state of an individual tracker T_i at time t by $\theta_{i,t}$ and its observations by $z_{i,t}$. The state transition model is:

$$\theta_{i,t} = \theta_{i,t-1} + N, \tag{1}$$

where N is the noise subscribing to a Normal distribution. Please note that the motion state itself is characterized by three variables: x, y for translation and ϕ for rotation.

The algorithm approximates the posterior distribution $p(\theta_{i,t}|z_{i,1:t})$ via a set of weighted particles $S_{i,t} = \{\theta^r_{i,t}, \omega^r_{i,t}\}^J_{r=1}$, where $\sum^J_{r=1} \omega^r_{i,t} = 1$; $S_{i,t}$ is properly weighted with respect to $p(\theta_{i,t}|z_{i,1:t})$. Then, we use the maximum a posteriori (MAP) estimate to determine the state of the tracker:

$$\hat{\theta}_{i,t} = arg \max_{\theta_{i,t}} p(\theta_{i,t}|z_{1:t}) \approx arg \max_{\theta_{i,t}} \omega^r_{i,t}. \tag{2}$$

The weight values of the particles are proportional to the posterior probability:

$$\omega_{i,t} \propto p(\theta_{i,t}|z_{i,t}). \tag{3}$$

In our implementation, the particle weight is computed as the correlation coefficient of the sampled region of interest (ROI) with an appearance template. The appearance template is composed of intensity values inside the ROI. The choice of pixel-based template ensures generality, rendering the method applicable beyond the MRI modality. Also, intensity blocks are computationally efficient. We adopt the spatio-temporal matte (STM) template described in [9]. The strong point of STM is that updating is based both on pixel dependence (spatial smoothness) and temporal dependence (temporal smoothness).

2.2 Collaborative Tracker Network ($n > 1$)

Particle filter trackers trade sophistication for generality and efficiency. This approach works well in simple motion scenarios, but may reduce robustness in challenging ones. By forming a collaborative network of particle filter trackers we aim to increase sophistication without sacrificing generality and efficiency. Figure 1(a) shows as an example a 3×3 tracker network ($n = 9$), where each tracker is assumed to interact with its neighbors. After having all tracker states computed via the corresponding particle filters, a survivor group is formed consisting of all the well performing trackers. For each individual tracker T_i, a decision is made whether to include it or not into the survivor group. The adjacent trackers $\{T_j, ..., T_m\}$ provide evidence to make this decision.

The effects of the adjacent trackers on tracker T_i are modeled via a Bayesian network (Figure 1(b)). $\hat{\Theta}_{i,t} = \{\hat{\theta}_{i,t}, \hat{\theta}_{j,t}, ..., \hat{\theta}_{m,t}\}$ and $Z_{i,t} = \{z_{i,t}, z_{j,t}, ..., z_{m,t}\}$ are the estimated states and observations of T_i and its adjacent trackers. $W_{i,t}$ represents the event that tracker T_i is in the survivor group at time t. $G_{i,t-1}$ is the Bayesian network at time $t-1$, whose probability $p(G_{i,t-1})$ is known at time t. The arrows in Figure 1(b) indicate dependency relationships. Two underlying assumptions for this Bayesian network are:

- An individual tracker is likely to have motion similar to its adjacent trackers.
- A tracker included in the survivor group at the previous time step is likely to be in the current survivor group.

The joint probability of the Bayesian network $G_{i,t}$ is:

$$p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t}) \propto p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t}) \prod_k p(\hat{\theta}_{k,t}|z_{k,t}), \tag{4}$$

where k identifies each of the trackers in the Bayesian Network and $p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t})$ is a probability function of $W_{i,t}$.

Given $G_{i,t-1}$ and $\hat{\Theta}_{i,t}$ as parameters, the function p is defined as:

$$p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t}) \propto p(G_{i,t-1}) \prod_{k} N(\hat{\theta}_{i,t}|\hat{\theta}_{k,t}, \sigma^2), \tag{5}$$

where $p(G_{i,t-1})$ is computed at time $t-1$ and is known at time t. $N(\hat{\theta}_{i,t}|\hat{\theta}_{k,t}, \sigma^2)$ is the probability density of $\hat{\theta}_{i,t}$ on the Normal distribution centered at $\hat{\theta}_{k,t}$ with variance σ^2. According to Equation (3) each tracker $p(\hat{\theta}_{k,t}|z_{k,t})$ in Equation (4) is proportional to the particle weight of the estimated tracker state. This equals to the normalized highest matching score among all the particles. Thus, the conditional probability $p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})$ on the left hand side of Equation (4) can be easily computed.

$p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})$ serves as evidence in deciding whether to include or not tracker T_i in the survivor group. If the evidence exceeds a minimum threshold, then it is included in the survivor group, otherwise it is excluded.

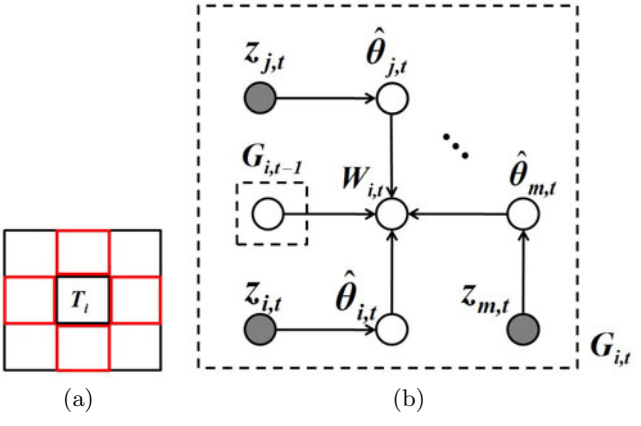

Fig. 1. (a) Layout of tracker network ($n = 9$). (b) Bayesian network for tracker T_i.

The overall motion state is determined by the trackers in the survivor group and computed as follows:

$$\hat{\theta}_{overall} = \frac{1}{\sum_{i=1}^{|W|} \beta_i} \sum_{i=1}^{|W|} \hat{\theta}_i \beta_i, \tag{6}$$

where $|W|$ is the cardinality of the survivor group and β_i is the impact factor of each linked tracker. The latter is determined from:

$$\beta_i = \frac{p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})}{D_{i,t}}, \tag{7}$$

where the enumerator is the conditional probability computed in Equation (4) and $D_{i,t}$ is the Euclidean distance from the point of interest (landmark) to the center of the tracker. Trackers closer to the landmark point weigh more on motion estimation with respect to those that are further away. The overall state is used to determine the new position of the landmark point as well as relocate the failed trackers.

3 Experimentation

3.1 Experimental Design

Data were acquired with a 1.5T Espree Siemens MRI scanner. The collected cine sets include short and long axis views on normal volunteers ($N = 2$) with a true fast imaging with steady-state precession (TrueFISP) sequence (TR/TE/a $= 60.3ms/1.4ms/80^o$; slice thickness $= 6mm$; acquisition matrix $= 256 \times 256$). We are interested in tracking anatomical landmarks on the heart, as a way to close the feedback control loop in MRI-guided robotic surgery on the valve. The interventional tool should be constantly adjusted with respect to the anatomical points to avoid harming critical structures of the beating heart. Specifically, there are four landmarks of interest (see Figure 2):

- Apex (A): The apical point of the left ventricle selected in long axis view.
- Medium (M): The center of the left ventricle at a basal level, i.e., just below the valves, in long axis view.
- Valve (V): The center of the entrance of the aortic valve annulus in long axis view.
- Centroid (C): The center of the left ventricle in short axis view.

Tracking experiments were performed on a set of 16 MRI cine sequences. Each sequence had 25 heart phases (frames) and total duration of 1 sec (approximately one heart-beat). Each of the 16 sequences had one landmark that belonged to one of the above categories. The goal for the experiments was to compare the speed, accuracy, and robustness of different tracking network configurations.

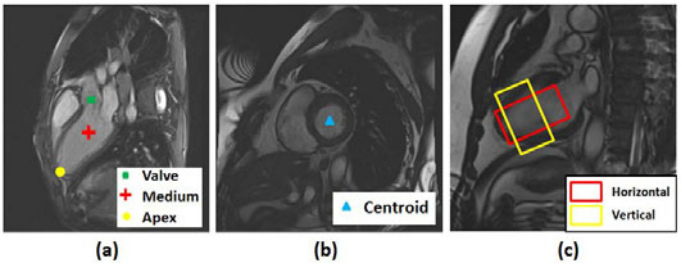

Fig. 2. (a)-(b)Anatomical landmarks of interest. (c) Tracker orientations: Horizontal (along the elongation axis of the moving part) and Vertical (across the elongation axis of the moving part).

These configurations featured different number, size, and orientation of trackers. Specifically, the following cardinalities and sizes were considered:

- Small single tracker (S1): One small tracker that tracks only the moving region around/near the landmark.
- Big single tracker (B1): One big tracker that in addition to the moving region includes some static surrounding structures.
- Collaborative trackers (B4): A 2×2 collaborative tracker network ($n = 4$) that has the same initialization as B1.

Each of these trackers can be applied in either horizontal (*H*) or vertical (*V*) orientation (Figure 2 (c)), leading to the following 6 configurations: *H*S1, *H*B1, *H*B4, *V*S1, *V*B1, and *V*B4. For the Centroid point, we make use of a square (instead of rectangular) tracker because the shape of the left ventricle that encloses, is approximately circular (instead of elliptical) in the short axis view. Thus, for this landmark type only, we have 3 instead of 6 configurations: S1, B1, and B4.

Ground-truth was manually labeled by an expert. The labeled points were recorded as time series of the coordinates of the relevant anatomical landmarks $\{x_g, y_g\}_t$, where t is the index of the heart phase ($1 \leq t \leq 25$). Then, all tracking configurations (6 for Apex, Medium, Valve and 3 for Centroid) were applied on the cine sequences to obtain time series of tracked estimates $\{x, y\}_t$ for the relevant landmarks. The closer the tracked time series $\{x, y\}_t$ to the corresponding ground-truth time series $\{x_g, y_g\}_t$, the better.

3.2 Experimental Results

In terms of speed, all proposed tracker configurations are computationally light and achieve real-time performance (25-40 fps) on a standard PC. In terms of accuracy, we use the Euclidean (L2) distance metric to measure the distance between the tracking results and the ground-truth values for all 16 cine sequences. We examine the following two questions: (a) For each landmark (Apex, Medium, Valve, or Centroid), which tracking configuration performs the best? (b) Of the three tracking configurations (S1, B1, or B4), which is the best overall?

In the table of Figure 3, we provide the L2 distances that each tacking configuration (columns) achieves on every cine sequence (rows). Cine sequences are named as c#:[Landmark Designator], where c# stands for the cine sequence code number (c1-c10), while the Landmark Designator is A for Apex, M for Medium, V for Valve, and C for Centroid. For left and right views in Apex/Valve the designator becomes L-A or L-V and R-A or R-V correspondingly.

The cells with red numbers in the table indicate the configurations that achieve the minimum L2 distance for each cine sequence. We observe that the network of collaborative trackers ($n = 4$) clearly outperforms the single trackers ($n = 1$). This is true for all types of landmarks. Only in the case of Apex, the easiest of the landmarks, the single trackers appear to be somewhat competitive. Indeed, the Apex landmark point features the least amount of motion and typically sits on a well contrasted tissue area. The big single trackers, have better contrasting support but some complex tissue movements from outlying areas

introduce ambiguity at times. The collaborative trackers distribute the risks and advantages. Therefore, failures are locally isolated and overall success is achieved through optimal probabilistic reasoning (Bayesian network). Another interesting observation is that the horizontal configurations fare a lot better than the vertical ones. This is probably to be expected, as the horizontal configuration is along the major motion axis, which is what the tracker strives to capture in the first place. The poor performance of the vertical configurations (which are limit cases), indicates that is not a good idea to orient rectangular trackers away from the major motion axis of the tissue.

		L2 Distance					
		*H*S1	*H*B1	*H*B4	*V*S1	*V*B1	*V*B4
Apex	c1:A	4.98	6.18	4.16	4.53	4.31	4.99
	c2:A	1.68	7.01	2.41	2.33	2.20	2.66
	c3:A	5.84	8.47	8.76	7.20	8.18	7.48
	c4:L-A	7.18	9.37	7.03	—	—	—
	c5:R-A	7.00	5.62	5.23	—	—	—
Medium	c2:M	30.74	5.25	3.91	6.14	4.20	4.08
Valve	c1:V	3.18	5.53	2.60	5.40	3.78	3.99
	c2:V	22.39	3.92	3.83	4.58	4.35	3.42
	c4:L-V	2.86	2.59	2.46	4.39	1.93	1.82
	c4:R-V	3.53	4.58	3.57	3.95	4.26	3.73
		S1	B1	B4			
Centroid	c5:C	16.11	2.39	1.65			
	c6:C	13.54	2.56	2.60			
	c7:C	31.87	3.07	3.05			
	c8:C	5.23	2.15	2.08			
	c9:C	3.59	3.09	1.58			
	c10:C	2.96	1.94	1.38			

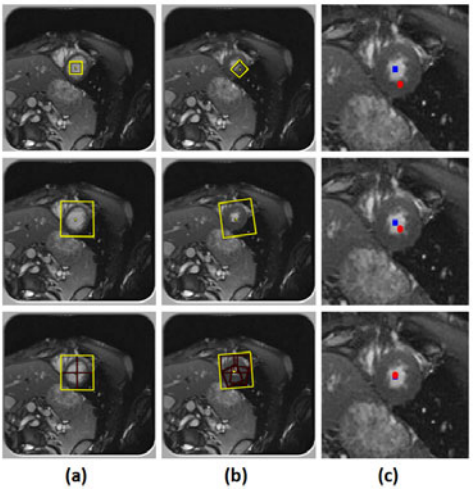

Fig. 3. Left: L2-based tracking performance. Right: Centroid tracking. (a) Initialization; (b) Ventricular contraction; (c) Ground-truth(blue) versus tracking results(red).

The right panel of Figure 3 shows annotated results from the tracking of the Centroid point for all three tracking configurations. As it is evident in the third column of images, the collaborative tracker configuration outperforms the single trackers; the reported landmark completely coincides with ground-truth. All the annotated experimental results can be found at `http://ourpapers.info/miccai10-mri`.

4 Conclusion

We have presented a collaborative tracking algorithm that can handle robustly heart motion as appearance changes (due to blood) in MRI cine sequences. The algorithm provides real-time information about landmark points through which robotic interventional tools can be compliantly guided in future valve repair operations. The individual trackers in the algorithm are unimpressive particle

filter trackers. When combined, however, under a Bayesian network framework, they produce sophisticated behaviors without loosing efficiency. The framework is flexible enough and general enough to be applied beyond MRI intervention studies.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under grants No. #IIS-0812526 and #CNS-0932272. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

References

1. Woo, Y.J.: Robotic cardiac surgery. International Journal of Medical Robotics and Computer Aided Surgery 2(3), 225–232 (2006)
2. Saikus, C.E., Lederman, C.J.: Interventional cardiovascular magnetic resonance imaging: A new opportunity for image-guided interventions. JACC Cardiovascular Imaging 2, 1321–1331 (2009)
3. Tsekos, N.V., Khanicheh, A., Christoforou, E., Mavroidis, C.: Magnetic resonance-compatible robotic and mechatronics systems for image-guided interventions and rehabilitation: A review study. Annual Review of Biomedical Engineering 9, 351–387 (2007)
4. McVeigh, E.R., Guttman, M.A., Lederman, R.J., Li, M., Kocaturk, O., Hunt, T., Kozlov, S., Horvath, K.A.: Real-time interactive MRI-guided cardiac surgery: Aortic valve replacement using a direct apical approach. Magnetic Resonance Medicine 56(5), 958–964 (2006)
5. Dowsey, A.W., Keegan, J., Lerotic, M., Thom, S., Firmin, D., Yang, G.: Motion-compensated MR valve imaging with COMB tag tracking and super-resolution enhancement. Medical Image Analysis 11(5), 478–491 (2007)
6. Fradkin, M., Ciofolo, C., Mory, B., Hautvast, G., Breeuwer, M.: Comprehensive Segmentation of Cine Cardiac MR Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 178–185. Springer, Heidelberg (2008)
7. Yuen, S.G., Kesner, S.B., Vasilyev, N.V., Nido, P.J., Howe, R.D.: 3D Ultrasound-Guided Motion Compensation System for Beating Heart Mitral Valve Repair. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 711–719. Springer, Heidelberg (2008)
8. Isard, M., Blake, A.: Condensation - Conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)
9. Zhou, Y., Tsiamyrtzis, P., Pavlidis, I.: Tissue Tracking in Thermo-physiological Imagery through Spatio-temporal Smoothing. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 1092–1099. Springer, Heidelberg (2009)